UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.87/L.7
18 March 1994

ENGLISH ONLY

THIRTEENTH UNITED NATIONS REGIONAL
CARTOGRAPHIC CONFERENCE FOR ASIA
AND THE PACIFIC
Beijing, 9-18 May 1994
Item 5 (d) of the provisional agenda*

NEW TRENDS IN TECHNOLOGY, AND THEIR APPLICATIONS: DIGITAL DATABASES, GEOGRAPHICAL AND LAND INFORMATION SYSTEMS

The U.S. National Spatial Data Infrastructure: continuing evolution

Paper submitted by the United States of America**

* E/CONF.87/1.

** Prepared by Nancy Tosta, U.S. Geological Survey.

INTRODUCTION

In a report published in 1993, the Mapping Science Committee (MSC) (1993) of the National Research Council defined the National Spatial Data Infrastructure (NSDI) as the "means to assemble geographic information that describes the arrangement and attributes of features and phenomena on the Earth. The infrastructure includes the materials, technology, and people necessary to acquire, process, store, and distribute such information to meet a wide variety of needs."

The term "spatial data," as used in the concept of the NSDI, refers to data that can be registered to a location on the surface of earth by a georeferencing system, such as latitude/longitude or through indirect referencing such as street addresses. These data are called geospatial data. The amount of geospatial data in the United States and the world is increasing at an exponential rate because of the proliferation of electronic recording instruments such as those carried on satellites or onboard ships, because of the use of geographic information systems (GIS) and global positioning systems (GPS), and because of the availability of high resolution, high speed scanning equipment. Such collection efforts have been spurred by the increasing number of relatively low cost, reliable computers and their ability to process, analyze, and display vast amounts of data.

Examples of geospatial data are (1) raster data such as satellite images, grids of topographic or sounding data, side scan sonar data from the seafloor, and images of maps or charts that have been electronically scanned; and (2) vector data such as lines defining highways, land use, administrative and international boundaries, hydrology, and shorelines.

Concerns have grown in the United States about the limited availability of digital geospatial data that are accurate, current, and the appropriate resolution. Other concerns have focused on the large sums of money, estimated to be $8 - 10 billion globally, that are expended annually on geospatial data by governments and the private sector. Recently, the Office of Management and Budget (OMB) conducted a data survey and determined that more than $4 billion is spent annually by U.S. Federal agencies on the collection, management, and dissemination of geospatial data. The questions raised are: Where are all those data? Why are they so difficult to find? When found, why are they difficult to use? Can they be integrated? Who maintains them?

The interaction of data producers and users in the United States is not the same as that in other parts of the world. Data are treated differently by the U.S. Government than almost any other institution in the world. For the most part, geospatial data produced at taxpayer's expense to meet the needs of the Federal Government are made available to anyone for the cost of dissemination. Data are not copyrighted (with some exceptions) and are free for redistribution, repackaging, or reformatting by others. Data sharing, at minimal or no cost, is a common practice among thousands of organizations. At local levels of government in the United States and in most other parts of the world access to geospatial data is much more controlled through copyrights and full-cost recovery policies.

There are numerous debates between levels of goverments in the U.S. and between the U.S. and other nations about the potential benefits and costs of making digital geospatial data more easily available. Given limited resources, how can more data be made available? Should data be freely distributed or should fees be levied? Does controlling access and charging for data improve data quality and availability? Are resources too limited to "give" data away? As the global information infrastructure contributes to increasing availability of digital data, issues of ownership, intellectual property rights, and data fees, will become even more complex.

In many ways, U.S. Government policies of freedom of access to data, are creating a greater need for coordination. The spread of technology and ease of data access have created an environment in which thousands of organizations require and produce data for use, sale, and distribution. Numerous data sets are available but they differ in accuracy, completeness, currency, and cost. Users have a difficult time finding and sorting through data to determine what might best meet their needs. Data producers often duplicate each others' efforts. Better means of collecting, organizing, finding, and sharing data are required to maximize investments being made in geospatial data.

About three years ago, discussions began on the need for better national coordination and organization among data users and producers. The concept of an NSDI was proposed by the MSC, as mentioned above. The NSDI is conceived to be a set of policies, procedures, standards, organizations, and data that contribute to the ability to more efficiently make use of and conduct analyses on geospatial data. Such an infrastructure would serve to integrate geographic information about features on the Earth. In the Fall of 1993, the Clinton-Gore administration identified the development of the NSDI as a key initiative for helping to reinvent government. The Federal Geographic Data Committee (FGDC), mandated by the

OMB, was charged with the responsibility of raising funds and overseeing partnerships with State and local governments and the private sector to facilitate NSDI efforts nationwide. In recent months, Secretary of the Interior Bruce Babbitt has assumed the chairmanship of the FGDC, raising the visibility of the NSDI within Federal agencies and across the Nation.

NSDI ACTIONS

Four main areas of action have been initiated to assist in developing the NSDI. The first is identifying needed standards and overseeing their development and implementation. The second is producing a framework of basic digital geospatial data from which other data can be derived or to which other data might be registered. The third is producing critical national digital thematic data sets, and the fourth is the development and evolution of the National Geospatial Data Clearinghouse. These activities are not mutually exclusive.

Standards

Standards are considered onerous by many who feel that they are too time consuming and overly complex. But the benefits can be significant. If, for example, standards did not exist for traffic patterns and road signals, far more accidents would occur. In other areas, standards for clothing and shoe sizes, building materials and batteries, for example, decrease costs and promote interchangeability. This does not mean that everyone has to wear the same size or color, live in identical houses, or buy identical electronic devices. Standards do not have to be that stringent. Standards are generally meant to facilitate, not control or dictate.

Similarly for geospatial data, there is not one size, accuracy, or common representation of geography that will fit all. But commonalities exist, and for many applications, common geographic resolutions and representations will suffice. Budgets often dictate using data that already exist, rather than designing projects around unique requirements. If a variety of digital geospatial data producers can agree to certain standards for how data will be collected and formatted, the ability to use these digital geospatial data would be greatly improved. This concept is addressed further in the framework data section.

Various subcommittees of the FGDC are working on data content standards that will help define how to map and classify specific themes of data such as topography and bathymetry, soils, wetlands, cadastral, and transportation. This effort does not mean that there are not

unique requirements for certain characteristics of these themes. But underlying most of these themes will be some commonality in definitions of classes and characteristics that will work for the majority of geospatial data users. Many users about to collect digital geospatial data related to one of these themes would find it useful to have guidelines from agencies who have studied these mapping activities. The standards will evolve as new tools and technologies for data collection affect what can be collected cost effectively and as more information must be integrated for more complex spatial analyses.

Additionally, standards that facilitate finding and accessing data are critical in today's world of rapidly increasing volumes of information. Data dissemination requires consistent, planned data management practices, such as documentation and archiving. Data that are unavailable do not contribute to human knowledge. The electronic highways of the future will carry terabytes of geospatial data. The ability to sort and filter through all of these data, extracting those that are most fit for use, is a requirement of the NSDI.

The FGDC is working on documentation or metadata standards to provide context for any geospatial data set. Additionally, transfer standards have been developed to facilitate use of data in different software systems. The spatial metadata standard is in the final stages of testing prior to release as an FGDC standard. This standard will most likely be submitted to the National Institute of Standards and Technology for adoption as a Federal Information Processing Standard (FIPS). The Spatial Data Transfer Standard (SDTS) has already been adopted as a FIPS. As of February 1994, Federal agencies are required, at a minimum, to make their data available to outside organizations in an SDTS format and to specify compliance to the standard in procurement of services from private sector vendors.

Framework Data

The fact that many users have common needs for digital geospatial data sets is germane to the concept of framework data. As spatial analyses are conducted, numerous users depend on certain geographic representations, either to reference the collection or display of other data or as variables for analysis. The U.S. Geological Survey products, such as 1:24,000-scale quadrangle maps, have formed a foundation in the analog world for geospatial data collection activities and analyses. Similarly, certain sets of digital data can form a framework for GIS analyses. If standardized digital geospatial data at 1:24,000 scale (roughly an accuracy of $\pm$ 12 meters) were complete for the Nation, a framework would exist on which to build many other data sources. The only complete national data bases are, however, at less accurate resolutions and are not current. An example is the Census Bureau's 1:100,000-scale TIGER (Topologically Integrated Geographic Encoding and Referencing) data having an accuracy of $\pm$ 50 meters.

At present rates of public investment for digitizing geospatial data, a national data base at a resolution and complexity comparable to the 1:24,000-scale quadrangles will not exist for decades. However, numerous disparate organizations are producing high resolution data (greater than 1:24,000 scale) over small geographic areas. Assuming these data meet some minimal level of standards, if they could be identified, made accessible, and "linked" together, major steps toward development of a framework might be accomplished. Given the lack of trusted, consistent data, or lack of knowledge about the existence of data, many

(perhaps thousands) of users are independently digitizing or capturing data at a variety of resolutions, in ad hoc, nonstandardized, one-shot approaches. Other users are working with new technologies such as GPS and digital ortho imagery (both satellite and photography), in an effort to more accurately represent geography.

The FGDC has created a working group to discuss the value to the Nation of implementing a minimum framework of maximally useful digital geospatial data in the shortest time possible and to define the characteristics of a framework. These discussions could lead to various agencies working to produce minimum versions of certain digital geospatial data sets as fast as possible. These data sets would be of a nature and resolution required by most people in a given geographic area. For example, framework data sets might include digital ortho imagery, elevation, and vector representations of transportation, shorelines, hydrology, or boundaries, at resolutions that would vary by geography, with higher resolution in urban areas and lower resolution in rural areas. Basic spatial representations of these themes would form the framework, such as delineations of transportation "pathways." More details about these "pathways" - for example, whether a pathway is a dirt road or 10-lane highway or a railroad - might be derived from other sources and referenced to the framework. The goal is to create a standardized, "trusted" geospatial framework within a short time to eliminate redundancy and duplication in data collection efforts and to minimize the difficulties of data integration through some common spatial denominators.

Many users are primarily interested in application-specific data, but large amounts of time are spent trying to find, collect, and organize basic framework information about roads or boundaries, for example. The FGDC is attempting to provide an organizational structure within which debates about how to approach these kinds of issues can be held and standards can be defined. If there could be more coordination in the collection of data to meet multiple users' needs, duplication could be minimized and more sophisticated data inventories over the same geographic area could be produced based on available funding.

Thematic Data Sets

The thematic data set component of the NSDI concerns those data that are required with some consistency on a national basis, but do not form the foundation for other data or are not necessarily required by most users, as in the case of framework data. The ability to collect or display thematic data often depends on the availability of framework data. Example thematic data sets are soils, land cover, and wetlands. Although these data may not be required by most users, there can be value in coordination of data collection practices. For example, there are national programs to protect wetlands which rely on the ability to accurately assess the extent and location of wetlands in a consistent fashion. Many local programs, however, also have wetlands assessment and management requirements, often demanding more detail than national efforts, so data collection and classification approaches vary. Different wetland classifications, criteria, and data collection techniques result in incompatible data sets, an inability to integrate data, and ultimately different estimates on the extent and condition of wetlands. Efforts are underway within the Wetlands Subcommittee of the FGDC to reconcile these discrepancies. The ability to aggregate a variety of data sets, through the use of agreed-upon standards, would enhance the availability of current data and provide data that are consistent and useful in multiple contexts. Expecting that a single

agency can compile and maintain data for the entire Nation within a meaningful timeframe is no longer reasonable. Changing the current situation requires standards, close coordination, data sharing, open data access policies, and trust in the data production efforts of other agencies.

National Geospatial Data Clearinghouse

The last, but hardly least important, component of NSDI is the clearinghouse. This is an effort to use the existing and rapidly expanding information infrastructure to facilitate access to geospatial data. This clearinghouse is distributed in nature, meaning there is not a centralized "warehouse" of geospatial data. Thousands, potentially millions, of data holdings will be cataloged and accessed across electronic networks.

There are three elements to the operation of the clearinghouse. The first is the FGDC spatial metadata standard. Data producers nationwide, at all levels of government and within the private sector, will be encouraged to document or describe, with the spatial metadata standard, any data they produce. This documentation will then be made electronically accessible on Internet. The Internet is the second element. Internet has become the most pervasive electronic network in the world, with an estimated 15 million users, and growth rates of 20 percent per month. This network of networks will continue to evolve and change as the private sector plays a larger role in the information infrastructure. Federal agencies, as well as other spatial data producers, are being encouraged to establish Internet connections. The third element of the clearinghouse is software tools for searching and querying data on the network. The FGDC has been testing wide area information servers (WAIS) software, which is built on a library standard (Z39.50) for searching and querying. WAIS was originally designed to conduct text searches, but has been enhanced to provide the ability to conduct geographic searches based on specifying the coordinates that cover an area of interest. New tools are continually being developed and offered to the public over the Internet. Recently, Mosaic, which provides the ability to build hypertext links between sources of data, has gained great popularity as a user interface for "surfing" the net. These three elements: metadata, Internet, and searching tools, together provide the ability to electronically document, serve, search for, browse, and access geospatial data that vary in scope, context, format, detail, and location.

In 1994, training and workshops about using the metadata standard, serving data, and using search and query tools are being offered by the FGDC. The success of the clearinghouse is dependent on participation. Telephones would not be particularly useful or deemed a success if few people used them. The clearinghouse is dependent on thousands of individuals modifying their behavior to more effectively manage their data, and doing so in a coordinated manner.

NSDI PARTNERSHIPS

Underlying all of these components of the NSDI is the need for creative partnerships. The NSDI builds on existing and yet-to-be formed institutions and relationships to facilitate the ability of the geospatial data community to share information and to develop and maintain

standard data sets. Progress and success of the NSDI will depend on participation from all levels of government, the private sector, and academia from all regions of the country.

In May 1993, the FGDC and various Federal agencies; GIS hardware, software and integration firms; and the major GIS professional societies in the United States sponsored the first National Geo-Data Policy Forum (1993). Individuals from government agencies, private corporations, and academia examined policies related to the NSDI and debated appropriate means of interaction to further the NSDI. No consensus exists on issues such as data ownership, privacy, copyrights, and data production responsibilities. The forum, however, provided the opportunity to identify issues and conduct constructive debates. The forum is likely to become a biennial event, with the next one scheduled for May 1995.

The forum was followed by public meetings held in conjunction with such major industry events as the Urban and Regional Information Systems Association annual conference, and the GIS/LIS conference. Each of these events provided a means to debate issues from a variety of public and private perspectives. Agreement exists that the policies for developing the NSDI must respond to the changing roles of the private and public sectors, must identify more creative ways for various communities to work together, must remove impediments to cooperation, while at the same time minimizing bureaucracy.

Within the Federal Government, support for the NSDI continues to grow. The Clinton administration endorses the vision of open access to public information and shared responsibilities in the development and use of that information. Vice President Gore's National Performance Review recommended the establishment of the NSDI as a national initiative to be based on partnerships with State and local governments and the private sector. The characteristics of a functioning NSDI - integration of different disciplines and efforts, along with a cooperative community spirit - are the same as those needed for government and society in dealing effectively with the challenges of the 21st century.

Opportunities for constructive communication and cooperation are needed. Partnerships work only if agencies, governments, and private companies have opportunities to interact, identify common concerns, and define needed products. The FGDC is conducting discussions at most of the major U.S. GIS conferences in 1994 and is working to optimize the various committee structures within FGDC to include non-Federal representation as appropriate. Additionally, State and regional councils and forums that consist of Federal, State, and local government and private and academic interests are being encouraged to facilitate dialogue on data needs and options for collecting and sharing information over areas of common geography. Such issues as the habitat concerns of the spotted owl or responding to the 1993 Mississippi River flood are geographically specific and involve many different institutions. More interactions to cooperatively build and manage shared data sets to respond to these geographically based concerns must be fostered.

All of these activities are very dependent on improved communication. At the Geo-Data Policy Forum, many people asked, "Who's in charge?" Perhaps this has become a world where that is no longer a relevant question. The interaction of computers and telecommunications is changing traditional philosophies on many things, but particularly

organizational hierarchies, economies, and information ownership. The increasing complexity of environmental, economic, and social problems on a global basis is forcing a reexamination of the interactions among individuals, governments, other sectors, and other nations. New ways to interact must be found to build new institutions and identify new leaders, all in a world where no single organization or institution is in charge. Cooperation and coordination based on respect and consensus building, rather than on dictating, demanding, or controlling, are critical to the success of the NSDI.

CONCLUSION

The concept of an NSDI has gained far-reaching acceptance within the geospatial data community. The FGDC has identified a number of actions that will assist in evolving the NSDI, but progress is dependent on the participation of thousands of disparate and separate individuals, organizations, and institutions. The ability to agree on and implement standards, to define commonly needed data sets, to share information, to communicate electronically, and to trust other agencies, are critical steps in building a successful NSDI.

REFERENCES

Federal Geographic Data Committee (1994). January 25 draft of *The 1994 Plan for the National Spatial Data Infrastructure: Building the Foundation of an Information Based Society*, Washington, D.C., 14 p. + Appendices.

Mapping Science Committee (1993). *Toward a Coordinated Spatial Data Infrastructure for the Nation*, National Academy Press. Washington, D.C., 171 p.

National Geo-Data Policy Forum Board of Directors (1993). *National Geo-Data Policy Forum Summary Report*, Association of American Geographers, Washington, D.C., 26 p.

Tosta, Nancy (1994). "The National Spatial Data Infrastructure: Where Are We Now?," *Geo Info Systems*, Vol. 4, No. 1, pp. 25-28.